1. Model:

For this week's prediction task, I chose a Random Forest Classifier model and tested it with three variations involving different configurations of ‘n_estimators’ and ‘max_depth’. This choice is sensible because Random Forest is known for its robustness and ability to handle both linear and non-linear relationships. It's also less likely to overfit compared to some other models.

2. Complexity of the Modeling Approach:

a. Base Model: The base model uses default parameters, serving as a benchmark.

b. Variations: Each variation adjusts ‘n_estimators’and ‘max_depth’, the maximum depth of each tree.

- Variation 1: A smaller number of trees without depth limitation.
- Variation 2: A moderate number of trees with restricted depth.
- Variation 3: A larger number of trees but with a shallow depth.

3. Hyperparameters Evaluated:

‘n_estimators’: Influences the number of trees in the forest. More trees can improve accuracy but also increase computational cost and risk of overfitting.

‘max_depth’: Controls the depth of each tree. Deeper trees can capture more complex patterns but might lead to overfitting.

4. Model Performance Metrics

	Variation	n_estimators	max_depth	AUC	AUC_Train
0	Base	100	NaN	0.854701	1.000000
1	Variation 1	10	NaN	0.831763	0.989448
2	Variation 2	100	10.0	0.814194	0.817013
3	Variation 3	500	3.0	0.683573	0.685281

a. Number of Trees (n_estimators):

The base model and Variation 2 use a moderate number of trees (100), while Variation 1 reduces this to just 10 trees and Variation 3 increases it to 500. Typically, increasing the number of trees can provide more robust predictions, but with diminishing returns as the number of trees increases.

b. Tree Depth (max_depth):

The depth of the tree can be indicative of the model's complexity. Deep trees might fit the training data very well, leading to higher training scores but might overfit and not generalize well to unseen data, leading to lower validation scores. This is evident in the table: The Base and Variation 1 models, which have unlimited depth, achieve perfect or near-perfect training AUCs, hinting at potential overfitting. Variation 3, with the shallowest trees, achieves the lowest AUC scores both in training and validation. This suggests potential underfitting as the model is too simple.

5. Best Model Selection and Analysis:

Based on the results, the Base Model is the top-performing model, with the highest AUC on the validation set. While it's common to consider more complex models with hyperparameter tuning, the results emphasize that default configurations of certain can sometimes provide robust performance, especially when they are well-calibrated for general use cases. In the context of this data, the base model's ability to grow trees without depth restrictions enabled it to capture the necessary patterns effectively, thereby offering the most reliable predictions among the variations. This underlines the importance of not only experimenting with various model configurations but also the value of benchmarking against default settings.